Diet and Serum Cholesterol

Do Zero Correlations Negate the Relationship?

David R. Jacobs, Jr., Joseph T. Anderson and Henry Blackburn


The confounding that results from the uncontrolled conditions under which most epidemiologic observations are made is sufficient to undermine their validity with respect to investigation of the relationship between diet and serum cholesterol. In this paper, the authors show, using both a mathematical model and referring to empirical data, that if certain variances are sufficiently great, even when there is cause and effect, correlation coefficients close to zero would be expected from the actual data of a cross-sectional study. Cross-sectional designs are therefore not suitable for studying this relationship.

biometry; cholesterol; cholesterol, dietary; coronary disease; dietary fats; nutrition surveys

The interest in serum cholesterol rests heavily on epidemiologic findings that people with elevated serum cholesterol have an elevated risk of having a heart attack. The findings of studies using free living volunteers throughout the world (1-12) are strongly complemented by those of animal studies, in which diet has been shown to cause atherosclerosis to develop (13-15) or recede (16, 17), and by autopsy studies of human beings (18). There is interest in diet because of controlled (19-26) and free-living (27) dietary experiments and clinical experience which indicate that reduction in the daily intake of dietary lipids will reduce serum cholesterol. Investigations on a cross-sectional population base have been made in several studies to see to what extent the relationship between diet and serum cholesterol carries over to the total populations. Both the Framingham Study (28) and the Tecumseh Study (29, 30) used careful dietary data collection methods. Correlations were examined between dietary lipid intake and serum cholesterol measured at the time the dietary history was taken. Zero or near zero correlations were found between the various components of diet and serum cholesterol level. Nichols et al. (29), in reporting these results from Tecumseh, state that the zero correlations "provide evidence that other factors besides fat intake are determinants of cholesterol levels among the general public" and further that "from the findings in this study one may infer that weight reduction should be the initial intervention for control of hyperlipidemia in the general population." The statements are correct.
that there may be other determinants of cholesterol level besides lipid intake (31–33) and that weight reduction may be a promising technique related to reduction of hyperlipidemia (27, table XII.28). However, the zero correlations do not provide evidence for these statements. Therefore, one of the major purposes of this paper is to point out the consistency of the findings by these cross-sectional studies of zero correlations with the results of controlled dietary experiments. While the concept that uncontrolled variation and unmeasured variables may obscure relationships between variables is well known, its importance in the understanding of diet and serum cholesterol seems to have been underestimated. It is concluded that in cross-sectional designs in the Tecumseh and Framingham studies, zero correlations do not negate the relationship.

**Controlled dietary experiments**

The controlled experiments which have found a relationship between intake of dietary lipids and serum cholesterol are elegant and persuasive. Keys et al. (19–22) on the basis of such experiments have developed a formula for a function, the change in which predicts the effect of dietary modification on serum cholesterol. Others have made minor refinements in the formula (34). The form in current use (35) is:

$$
\phi = 1.26 (2S - P) + 1.5 (1000 C/E)^{1/4},
$$

where $S$ = per cent of calories as saturated fatty acids, 12–16 carbon chain lengths; $P$ = per cent of calories as polyunsaturated fatty acids; and $1000 C/E = \text{mg dietary cholesterol}/1000 \text{kcal}$.

This function involves: 1) saturated fatty acids, which increase serum cholesterol; 2) polyunsaturated fatty acids, which decrease serum cholesterol at half the rate; and 3) dietary cholesterol, which increases serum cholesterol according to a square root. The function is designed so that for a person with an initially average serum cholesterol, change of one unit of $\phi$ should result in a change of 1 mg/dl in serum cholesterol. Actually, individual responsiveness seems to be greater when serum cholesterol is initially higher (36).

One such experiment is illustrated here (24). Table 1 shows the design: Thirty-eight subjects were divided into four subgroups. Each individual and each subgroup went through six dietary periods, each of four weeks duration, eating highly controlled diets. The duration was set at four weeks to give a period long enough for serum cholesterol levels to stabilize under a changed regime (26), and thus to avoid a carryover effect. The diets were formed by taking a basic diet of 1790 kcal and adding supplements of various kinds of fat or carbohydrate providing 680 kcal to this base. In the mnemonic designations of the diet, O refers to olive oil, S to safflower oil, P to palm oil, and the numbers refer to grams of oil mixture in the daily supplement. In diet CHO the

### Table 1

*Design of a controlled dietary experiment showing relationship between $\Delta$ serum cholesterol and $\Delta\phi$ (Design: symmetric Latin square ($N = 38$))*

<table>
<thead>
<tr>
<th>Group</th>
<th>Four-week diet periods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>W</td>
<td>Butter</td>
</tr>
<tr>
<td>X</td>
<td>Butter</td>
</tr>
<tr>
<td>Y</td>
<td>Butter</td>
</tr>
<tr>
<td>Z</td>
<td>Butter</td>
</tr>
</tbody>
</table>
supplement was sugar and jelly and in diet "Butter" the supplement was butter. Periods 2, 3, 4, and 5 constitute a symmetric Latin square.

The five diets considered in this experiment varied according to percentage of calories as fat and according to level of $\phi$ (table 2). The butter diet eaten at the beginning and the end had 31 per cent of calories as fat and $\phi = 47$. The other four diets had nearly constant $\phi$ of 13–15 and percentage of calories as fat ranging from 5 through 30. The four experimental diets based on carbohydrates or various mixtures of oil within natural diets differ widely in percentage of calories from fat, but differ hardly at all in terms of the level of $\phi$. The range of average serum cholesterol observed on these four diets is 221–228 mg/dl, with standard error approximately 6.5 mg/dl in each case. As predicted by $\phi$, this small range is consistent with the notion that serum cholesterol is unaffected by change in percentage of calories due to fat, provided the value of the function $2S - P$ is constant. That serum cholesterol is affected by change in $\phi$ is shown by the large difference between serum cholesterol on the butter diet and on the other four diets. Consistency is shown by the similarity of serum cholesterol levels on the butter diet, both before and after changes associated with eating the other diets. This experiment controlling many of the sources of variation discussed later is an elegant and strong validation of $\phi$ as a predictor of change in serum cholesterol. Strictly speaking, these results relate only to men who are in caloric balance and in a situation in which fats are substituted for refined sugar or monounsaturated fatty acids. The experiment does not rule out the possibility that other dietary factors also affect serum cholesterol.

A MATHEMATICAL MODEL

Assuming that the underlying relationship in cross-sectional observations of diet and serum cholesterol is linear with perfect correlation ($p_{x,y} = 1$), we can examine what to expect in a real situation when serum cholesterol and diet both are measurable, if only inexactly.

Let serum cholesterol correspond to $Y$ and the diet function predicting serum cholesterol correspond to $X$. $Y$ and $X$ denote parameter values that would be observed if measurements could be made without error. For a population of $N$ individuals, all of whom alike in cholesterol metabolism but eating varied diets, the model of the relationship (for some fixed parameters $a$ and $b$) is

$$Y_i = a + bX_i$$

$i = 1, \ldots, N$.
Exact equality holds because $\rho_{x,y} = 1$. The same relationship holds for the means, $X$ and $Y$, $\bar{Y} = a + b\bar{X}$.

The scales of $X$ and $Y$ can be changed by setting $x_i = b(X_i - \bar{X})$ and $y_i = Y_i - \bar{Y}$, giving the simple expression

$$y_i = x_i, \quad i = 1, \ldots, N. \quad (1)$$

For this expression $\bar{y} = \bar{x} = 0$, $\sigma_x^2 = \Sigma x^2/N$, $\sigma_y^2 = \Sigma y^2/N$, and $\rho_{x,y} = 1$. Summations here and below are intended for $i = 1, \ldots, N$. Variances $\sigma_x^2$ and $\sigma_y^2$ quantify differences between serum cholesterol levels of individuals and between food patterns of individuals, respectively. Here serum cholesterol levels and food patterns are assumed to be constant within each individual. In the particular case in which $X$ corresponds to $\phi$, $a = 164$ and $b = 1$ for a man with a typical cholesterol metabolism (35). Many of the men studied in controlled experiments seem to be typical, with $b$ between 0.8 and 1.2 (34, 36). Assuming that $b = 1$, the scale of $x$ is the same as that of $X$ and $\sigma_x^2 = \sigma_Y^2$. If $b \neq 1$, $\sigma_x^2 = b^2 \sigma_Y^2$.

In any real situation the observed values of $y$ (serum cholesterol) with diet fixed will be subject to variability relating to blood sampling and chemical analysis and to other variation in cholesterol levels both within and between individuals, unrelated to diet. The importance of the various sources varies and may depend on the situation. Many factors, such as use of medication, season, posture, blood drawing technique, age, race, and sex are routinely controlled. Improved chemical methodologies and standardization of blood drawing conditions provide satisfactory control (37). Spontaneous within-individual variability of serum cholesterol is relatively large. A standard deviation of 10 to 12 mg/dl over 2–6 weeks has been reported (38). For free-living individuals on apparently constant diets, a within-individual standard deviation of 15 to 20 mg/dl is more typical (38–41). This variability cannot be controlled because its sources are not well understood. Differences in intrinsic serum cholesterol can be dealt with using a design with multiple observations for each individual, discussed below. If the sum of these components of variation is designated by $e$, the model for observed serum cholesterol becomes

$$y_i' = x_i + e_i, \quad i = 1, \ldots, N. \quad (2)$$

In this expression $x$ and $e$ are assumed not to be correlated, and

$$\bar{y}' = \bar{x} = \bar{e} = 0, \quad \sigma_x^2 = \Sigma x^2/N, \quad \sigma_e^2 = \Sigma e^2/N.$$

By definition

$$\rho_{x,y'} = \rho_{x,y} \frac{\Sigma xy'}{\sqrt{\Sigma x^2} \sqrt{\Sigma y'^2}}.$$

Substitute $(x + e)$ for $y'$

$$\rho_{x,y'} = \frac{\Sigma x (x + e)}{\sqrt{\Sigma x^2} \sqrt{\Sigma (x + e)^2}}.$$

Since $x$ and $e$ are not correlated, $\Sigma xe = 0$

$$\rho_{x,y'} = \frac{\Sigma x^2}{\sqrt{\Sigma x^2} \sqrt{\Sigma x^2 + \Sigma e^2}} \quad (3)$$

$$= \frac{\sigma_x^2}{\sqrt{\sigma_x^2 + \sigma_e^2}}.$$

Thus the population value for the correlation coefficient is the square root of a ratio between variances. This shows that the coefficient of correlation is reduced from unity (attenuated) when the variability in serum cholesterol values is introduced. This argument is very similar to one used by Berkson (42). Note the algebraic equivalent to formula 3: $\sigma_e^2 = (1 - \rho_{x,y'}) (b^2 \sigma_X^2 + \sigma_e^2)$. Here $1 - \rho_{x,y'}$ is identified as the proportion of the total variance of $Y$ which represents variance away from the regression line.

Expressions 2 and 3 are oversimplified because actual cross-sectional data are subject not only to variability in $y'$ (serum
cholesterol) but also to variability in \( x \) (the diet function). These sources of variability include: 1) errors in identifying food items in the food table, 2) discrepancies between the food table value and the true composition of the food eaten, 3) errors in estimating quantities of food eaten, 4) errors in remembering what was eaten, and 5) differences between the food pattern of the observation period and the food pattern of the previous two to four weeks. Another similar source of variance is the variability in diets of components that influence serum cholesterol but which have not been included in the serum cholesterol predicting function \( x \) (lack of fit of the model). Of these sources, item 5), variation in day-to-day eating pattern, is probably the most significant. Control of this source of variation could be achieved by obtaining an accurate long-term food record; this maneuver may not be feasible in cross-sectional designs (43), but is achieved in feeding experiments. If we lump all these sources of variability under the designation \( \epsilon \) we can write for the observed value of the dietary function

\[
x'_i = x_i + \epsilon_i \quad i = 1, \ldots, N. \quad (4)
\]

By the argument leading to formula 3, but inserting \( x' = x + \epsilon \) for \( x \) in formula 3, it follows that the correlation between observed diet function and observed serum cholesterol for individuals is

\[
\rho_{x',y'} = \frac{\sigma^2_{x} - \sigma^2_{\epsilon}}{\sqrt{\sigma^2_{x} + \sigma^2_{\epsilon}} \cdot \sigma^2_{x} + \sigma^2_{\epsilon}}. \quad (5)
\]

This expression shows how the correlation between observed serum cholesterol and diet function is further diminished by the existence of variability of the values of diet function. This coefficient of correlation depends on differences between food patterns of individuals (\( \sigma^2_{x} \)), inexactness in the measurement of customary diet (\( \sigma^2_{\epsilon} \)), and inexactness in the measurement of average serum cholesterol (\( \sigma^2_{x} \)).

The above arguments pertain directly to the entire population of values. A cross-sectional sample would be expected to exhibit similar behavior. In a sample, \( r_{x',y'} \) would exhibit variability, but would estimate \( \rho_{x',y'} \).

It is possible to estimate the magnitude of these variances and to predict the attenuated correlations expected in cross-sectional data. The range of habitual values of \( \phi \) in the general population is probably 40 to 80, which would suggest a standard deviation for \( \phi \) or \( x \) of 10. That such a range is reasonable is suggested in part by considering the means and standard deviations for the dietary constituents in the National Diet-Heart Study report (27, table IX.3). If \( \sigma^2_{x} \) is smaller this would tend to further attenuate correlation coefficients. We assume \( \sigma^2_{x} = 100 \). The variance of \( \epsilon \) can be estimated similarly from the range of serum cholesterol values in groups eating identical diets. A typical range for serum cholesterol for a diet squad of this kind is 130 to 250 mg/dl. This indicates a standard deviation of 30 and variance, \( \sigma^2_{y} \), of 900, nine times as great as \( \sigma^2_{x} \). Accordingly, expression 3 has a value approximately \( \sqrt{1/1 + 9} \) or 0.316 (see table 3, row 1, column 7).

The variance of \( \epsilon \), that is, the variability associated with data for the diet function, includes as a major component the differences between the values for the diet of the day of observation and the diet of the period that is physiologically related to the observed serum cholesterol level, the average diet of the preceding three or four weeks. For a typical individual the daily saturated fatty acid range over a period of a few weeks may be from 10–22 per cent of total calories. Similarly, dietary cholesterol might range from 200–800 mg per day, and polyunsaturated fatty acids from 2–6 per cent of total calories. The extremes of \( \phi \) computed for such an individual would be 30 and 77. This suggests a standard devia-
Expected coefficient of correlation score, attenuated by the population variability of serum cholesterol and by the within-individual variability of diet

<table>
<thead>
<tr>
<th>Variance ratio, $\sigma_Y^2/\sigma_X^2$</th>
<th>0</th>
<th>2/3</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>9</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0.816</td>
<td>0.707</td>
<td>0.577</td>
<td>0.447</td>
<td>0.378</td>
<td>0.316</td>
<td>0.250</td>
</tr>
<tr>
<td>2/3</td>
<td>0.894</td>
<td>0.730</td>
<td>0.632</td>
<td>0.516</td>
<td>0.400</td>
<td>0.338</td>
<td>0.283</td>
<td>0.224</td>
</tr>
<tr>
<td>1</td>
<td>0.816</td>
<td>0.667</td>
<td>0.577</td>
<td>0.471</td>
<td>0.365</td>
<td>0.309</td>
<td>0.258</td>
<td>0.204</td>
</tr>
<tr>
<td>2</td>
<td>0.707</td>
<td>0.577</td>
<td>0.500</td>
<td>0.408</td>
<td>0.316</td>
<td>0.267</td>
<td>0.224</td>
<td>0.177</td>
</tr>
<tr>
<td>4</td>
<td>0.577</td>
<td>0.471</td>
<td>0.408</td>
<td>0.333</td>
<td>0.258</td>
<td>0.218</td>
<td>0.183</td>
<td>0.144</td>
</tr>
</tbody>
</table>

$\sigma_Y^2$ = Variance of serum cholesterol in single observations of each individual in the population.

$\sigma_X^2$ = Variance of dietary score in repeated observations of a typical individual over a period of 3 or 4 weeks.

$\sigma_Y^2$ = Variance of individual dietary scores in single observations of the average dietary score over 3 to 4 weeks for each individual in the population.

The "scrambling effect" of variability is simply but realistically depicted in figure 1, which shows schematically an underlying perfect linear relationship between dietary score, $\phi$, and predicted serum cholesterol, $Y$, and the effect of introducing typical variations into these values such as might be observed for five individuals. The five open circles in figure 1A conform to the equation $Y = \phi + 164$. The effect of variability on serum cholesterol, $e$, is represented by the vertical arrows. The values of $e$ chosen are $-1\frac{1}{2}$, $-3/4$, $+1/10$, $+3/4$, and $+1\frac{1}{2}$ standard deviations of serum cholesterol and the five $e$ values are randomly assigned to the five points. In a similar way the effect of variability on dietary score, $\epsilon$, is represented by the horizontal arrows. Typical sources of variation are discussed for $e$ before formula 2 and for $\epsilon$ before formula 4. In this calculation, 30 and 10 are used as standard deviations of serum cholesterol and dietary score, respectively. By applying these shifts to the points a new set of points is obtained that simulates the expected observations of five individuals. The new points are shown in figure 1B. It is obvious that the correlation has been obscured by the random variability. Properly estimating the relationship falls within the purview of classical experimental design (44), as suggested by the use of the symmetric Latin square in the controlled experiment (24), and as further discussed below.

### Eliminating Effects That Do Not Change With Time—Intrinsic Level of Serum Cholesterol

One source of variation in serum cholesterol data was described above as between individual variability due to intrinsic differences between persons. We use the term intrinsic level to designate the serum cholesterol level that would be attained by the given individual if he or she were to eat a diet with a score of $\phi = 0$. Such a diet might, for instance, have no dietary cholesterol and twice as much polyunsaturated fat as saturated...
A measure of difference in intrinsic level could equally well be obtained by comparing individuals on any reference diet with fixed $\phi$. It is well known that serum cholesterol level will differ from individual to individual on a diet with $\phi$ held constant (24). We conceive of this intrinsic level of serum cholesterol as remaining constant as time passes. This constancy can be utilized analyti-

If the intrinsic level were known for each individual then it could be utilized to adjust the serum cholesterol level. This adjusted level would increase the accuracy of a cross-sectional analysis. Such a design, however, does not lead to knowledge of the intrinsic level. In fact, the intrinsic level is glossed over in a cross-sectional design as part of the general variation, which is represented in equation 2 by the symbol $e$. If in the effort to establish a relationship between diet and serum cholesterol the experimental design were set up so that two observations were taken for each individual in contrasting dietary situations, then the difference between the two serum cholesterol levels would no longer involve the intrinsic level. While it would not be explicitly known in either the first or the second observation, its contribution to each observation would be the same for any given individual. Its effect would thus be eliminated in the subtraction. The final effect is a reduction in the variance of the estimate of the slope of the relationship. The use of contrasting diets tends to further reduce the variance of estimation.

In the simplest situation, assuming linearity of the intrinsic level, two observations of $y(y_1, y_2)$ and of $x(x_1, x_2)$ would be available for each individual. These observations would be related by the formulas $y_1 = x_1 + z$ and $y_2 = x_2 + z$, where $z$ is a source of error which remains constant within each individual as time changes. Then

$$y_2 - y_1 = (x_2 + z) - (x_1 + z) = x_2 - x_1$$

and in this simple situation the assumed linear relationship between $y$ and $x$ can be seen exactly. Carryover effects are assumed to be accounted for by sufficient time between first and second observations. Cross-sectional observation of

![Figure 1. A. The open circles show the underlying relationship that exists in five typical persons between diet score ($\phi$) and predicted serum cholesterol, $\phi + 164$. Vertical arrows represent random variations in serum cholesterol and horizontal arrows random variations in diet scores. The resulting positions at the heads of the horizontal arrows represent typical observations to be expected. B. The points are the positions in A after introduction of variations. The perfect correlation of the points in A has been almost completely destroyed.](image-url)
serum cholesterol and diet is more complicated but contains components of the simple situation. The symbols \( e \) in equation 2 and \( e \) in equation 4 represent the composite of a number of sources of variation. The intrinsic level is analogous to \( z \) in equation 6. One way of estimating how much of \( \sigma_e^2 \) is due to intrinsic serum cholesterol levels would be to examine the variance of serum cholesterol for people on an ad lib diet and then again for the same group on a diet with a fixed \( \phi \). The National Diet-Heart Study provides an approximation to this situation (27, table XI.3). For open centers for diet B the standard deviation of serum cholesterol at baseline was 35.7 mg/dl. After 12 weeks on the diet it was 34.2 mg/dl, which is 92 per cent of the original variance. In Faribault where the adherence to diet B was presumably better, the standard deviation was 36.8 mg/dl during the baseline period and after 12 weeks on the assigned diet it was 29.2 mg/dl. This is 63 per cent of the original variance. Other estimates of the portion of \( \sigma_e^2 \) due to intrinsic serum cholesterol level can be obtained from the same source; the conclusion remains that intrinsic level of serum cholesterol constitutes a significant portion of \( e \).

**AN APPLICATION OF THE MODEL TO A FREE-LIVING EXPERIMENTAL POPULATION**

Designs with multiple observations per person in contrasting dietary situations have been the backbone of controlled dietary experiments. Such designs control two important sources of variation: intrinsic level of serum cholesterol and lack of accurate knowledge of dietary composition for a period of several weeks. We have used the same principle in analyzing data from a free-living experimental population.

The men involved in this study were 91 of the first 103 participants who were found eligible for the Minnesota clinic of the Multiple Risk Factor Intervention Trial (MRFIT) (45, 46) on the basis of high risk of coronary heart disease as indicated by serum cholesterol level, blood pressure and cigarette smoking habit and who were assigned by a random process to the half of the participants who received "special care," i.e., intense intervention to change blood pressure and eating and cigarette smoking habits.

The intervention program in which these special care participants were involved included a series of 10 weekly classes and individual counseling by a physician, a nutritionist and a behavioral counselor. Values of the dietary score, \( \phi \), were compared before and after three to six weeks of intervention. For each man, the starting food record was a 24-hour recall of food eaten on one day, the day before he was randomly assigned to treatment (his third clinic visit). This recall was assisted by an experienced nutritionist and recorded by her. Food models were used to assist in estimating the size of each serving. In spite of sincere effort to make the record accurate, it is certain that some errors occurred. It is also clear that the record of a single day, even if completely accurate, does not precisely describe the habitual food pattern of an individual. The food record for the later period, after food habits had been influenced by the weekly classes, consisted of a three-day record prepared at home, day-by-day, by the participant and his wife. The record was reviewed by a nutritionist in consultation with the participant in an effort to include items easily overlooked, to get exact descriptions (such as the type of margarine, the kind of oil, etc.), and to verify serving sizes. It is expected that a three-day record is a more precise estimate of food habits than a one-day record. Making the record at the time the food is being eaten is expected to give a better list of what was consumed than recall made on the following day. Both of these factors tend to make the intervention food...
values more precise than the baseline values in terms of having a smaller intra-individual variance or a narrower confidence interval for the mean dietary score values typifying the individual. These dietary records were manually coded using the data of Agriculture Handbook Number 8 (47), supplemented by food cholesterol values from a magnetic tape record (Data Set 8-1-1, March 1972, Agricultural Research Service, Consumer and Food Economics Institute, Federal Center Building, Hyattsville, MD 20782). Serum cholesterol was observed before dietary change, $SC_1$, and two months later after extensive dietary change, $SC_2$. $\phi_1$ and $\phi_2$ were computed, based on corresponding dietary records. The results are consistent with the mathematical model presented. The correlation between the initial cholesterol level, $SC_1$, and the initial dietary characterization, $\phi_1$, a cross-sectional correlation, is equal to $-0.1$. Similarly and coincidentally, after the dietary change, the correlation between $SC_2$ and $\phi_2$ is $-0.1$. This is similar to the findings in the Framingham Study (28) and the Tecumseh Study (29, 30) using similar careful dietary coding techniques and similar food tables. On the other hand, change in serum cholesterol, $SC_1 - SC_2$ and change in $\phi_1 - \phi_2$, have a correlation coefficient among these 91 men of 0.4, highly statistically significant under the null hypothesis $\rho = 0$. The analysis of changes in diet and changes in serum cholesterol has partly eliminated a confounding factor: the intrinsic level of serum cholesterol. The correlation 0.4 is still considerably reduced from 1, however, presumably due to the remaining sources of variation. While predictability of serum cholesterol changes from diet changes is not great here, the consistency of this finding with the theory presented above is apparent. This finding belies the notion that diet and serum cholesterol are somehow related in groups but not in individuals.

**Discussion**

The question of discerning a relationship between diet and serum cholesterol has been considered in this paper. Various study designs are considered, including controlled feeding experiments, free-living dietary change experiments, and cross-sectional observation of large populations. A clear and strong relationship has been found between change in diet and change in serum cholesterol in the former two designs. In cross-sectional designs, however carefully carried out, correlation coefficients of approximately zero have been reported between dietary factors and serum cholesterol levels.

We have proposed a mathematical model to explain this apparent anomaly. A report is given from a small free-living dietary experiment which empirically supports the model. The results given here imply that the zero correlations found by various cross-sectional observational studies actually do not negate the evidence that diet has an effect on serum cholesterol. In statistical parlance, a cross-sectional study has near zero power for detecting such a relationship. It is expected that correlations near zero would arise between diet and serum cholesterol whether they are related or not. Thus cross-sectional observation of this relationship seems to be inappropriate.

While the observations in Framingham (28) and Tecumseh (29, 30) have been quite useful in delineating the relationship between serum cholesterol and coronary heart disease and in characterizing the types of food that total communities are eating, they are inappropriate to the question of whether or not serum cholesterol and diet are related. If the investigator knows a person’s customary dietary intake this will not tell the investigator the level of serum cholesterol of the person, even though, if that person modifies his diet to reduce his intake of dietary lipids, his serum cholesterol will go down.
from its present level. After dietary modification in a total population there will still be considerable differences between people in terms of their serum cholesterol levels and, therefore, in terms of their risk levels.

Controlled dietary experiments (19–26, 30–33, 35, 38) and the National Diet-Heart Study (27) have clearly shown that serum cholesterol decreases can be achieved in most individuals by appropriate dietary modification. If diet is appropriately changed, serum cholesterol will be reduced. The hypothesis that serum cholesterol level is directly related to the development of atherosclerosis implies that risk itself will be reduced considerably. The link of diet to coronary heart disease is presumably not direct but is through its effect on serum cholesterol. Since diet and serum cholesterol have a zero correlation cross-sectionally, a study of the relationship between diet and coronary heart disease incidence will suffer from the same difficulties as the study of diet and serum cholesterol. A corollary of the mathematical model here presented is that a correlation close to zero would likely be observed between diet and coronary heart disease incidence. For another disease in which diet is a cause agent, a prospective incidence study following a cross-sectional assessment of diet would presumably be valuable. An appropriate design for demonstrating or refuting diet and coronary heart disease incidence is a dietary change experiment.

REFERENCES